GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:11 ; Search time 8498.8 Seconds
(without alignments)
31.610 Million cell updates/sec

Title: US-09-851-670-4
Perfect score: 25
Sequence: 1 acagtagcagcaacagcatgagacc 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0
Maximum DB seq length: 60

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	16.2	64.8	58	10	AI759324	AI759324 EtESTea27
c	2	16	64.0	32	11	N40655	N40655 yw78b10.r1
c	3	16	64.0	55	11	BF346180	BF346180 602017646
	4	15.2	60.8	52	10	AA996148	AA996148 os14b11.s
	5	15	60.0	38	13	AQ025540	AQ025540 EP(X)1614
c	6	15	60.0	49	11	H66141	H66141 yu16e05.s1
	7	14.8	59.2	52	11	C20867	C20867 HUMGS000493
c	8	14.8	59.2	53	13	TA180E05P	AL474283 T. brucei
c	9	14.8	59.2	57	13	AZ616824	AZ616824 1M0446L15
	10	14.6	58.4	24	13	AZ779573	AZ779573 2M0016K09
c	11	14.6	58.4	27	13	AZ404206	AZ404206 1M0172I20
	12	14.6	58.4	35	13	AZ861400	AZ861400 2M0167O13
c	13	14.4	57.6	40	13	AZ832139	AZ832139 2M0112P14
c	14	14.4	57.6	51	13	AZ488023	AZ488023 1M0318J02
c	15	14.4	57.6	53	13	TA263C05P	AL483802 T. brucei
	16	14	56.0	46	10	AI884025	AI884025 fc72c11.x
	17	14	56.0	50	10	AU103544	AU103544 AU103544
c	18	14	56.0	50	10	AU104442	AU104442 AU104442
c	19	14	56.0	52	10	AA856040	AA856040 vw82a06.r
c	20	14	56.0	52	10	AA615079	AA615079 vo34b11.r
c	21	13.8	55.2	50	10	AU102281	AU102281 AU102281
c	22	13.8	55.2	50	10	AU105836	AU105836 AU105836
c	23	13.8	55.2	57	10	BE408921	BE408921 601303913
	24	13.6	54.4	28	10	AI755903	AI755903 EtESTea08
c	25	13.6	54.4	56	10	AA587506	AA587506 nn85e03.s
c	26	13.6	54.4	60	13	AZ651227	AZ651227 1M0521M13
	27	13.4	53.6	50	10	AU102654	AU102654 AU102654
c	28	13.4	53.6	52	10	BE321070	BE321070 NF033E03R
	29	13.4	53.6	56	13	AZ820546	AZ820546 2M0092J07
	30	13.2	52.8	21	13	AZ831993	AZ831993 2M0112M01
c	31	13.2	52.8	38	13	AZ971376	AZ971376 2M0244B21
c	32	13.2	52.8	39	13	AZ635338	AZ635338 1M0491E10
c	33	13.2	52.8	44	10	AV836720	AV836720 AV836720
c	34	13.2	52.8	50	10	AV833524	AV833524 AV833524
c	35	13	52.0	32	13	TA337F07Q	AL491668 T. brucei
c	36	13	52.0	37	13	AZ807121	AZ807121 2M0069I11
c	37	13	52.0	41	10	AA509356	AA509356 vg18a09.r
c	38	13	52.0	43	10	AA937113	AA937113 oc11h08.s
c	39	13	52.0	50	11	BF537767	BF537767 602054466
	40	13	52.0	53	13	AF179999	AF179999 AF179999
c	41	13	52.0	53	13	AF180000	AF180000 AF180000
c	42	13	52.0	54	11	BF163806	BF163806 601772312
	43	13	52.0	56	13	HSMC02B06	X88506 H.sapiens D
c	44	13	52.0	59	10	AW687832	AW687832 NF014A01R
	45	13	52.0	60	13	AZ623437	AZ623437 1M0461L08

ALIGNMENTS

RESULT 1
AI759324
LOCUS AI759324 58 bp mRNA EST 18-JAN-2000
DEFINITION EtESTea27b06.y1 Eimeria S5-2 Sporozoite stage Eimeria tenella cDNA 5', mRNA sequence.
ACCESSION AI759324
VERSION AI759324.1 GI:5174991
KEYWORDS EST.
SOURCE Eimeria tenella.
ORGANISM Eimeria tenella
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.
REFERENCE 1 (bases 1 to 58)
AUTHORS Liberator,P., Diaz,C., Tang,K., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Florence,N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R., Wilson,R. and Sibley,D.
TITLE WashU-Merck Eimeria tenella project
JOURNAL Unpublished (1999)
COMMENT Contact: David Sibley, Ph.D.
WashU-Merck Eimeria tenella project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
Trace considered overall poor quality
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .58
/organism="Eimeria tenella"

/strain="LS18"
/db_xref="taxon:5802"
/clone_lib="Eimeria S5-2 Sporozoite stage"
/dev_stage="Sporozoite"
/lab_host="SOLR E. coli"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E. tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on Sephacryl S500. cDNAs were digested with EcoRI/XhoI and cloned into lambda Zap II (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and SOLR cells (Stratagene). Insert sizes range from 1.2-2.9 kb."

BASE COUNT 19 a 17 c 20 g 2 t
ORIGIN

Query Match 64.8%; Score 16.2; DB 10; Length 58;
Best Local Similarity 85.7%; Pred. No. 1.1e+04;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 cagtagcagcaacagcatgag 22
||| ||||||| ||||| |||
Db 38 CAGCAGCAGCAGCAGCACGAG 58

RESULT 2
N40655/c
LOCUS N40655 32 bp mRNA EST 22-JAN-1996
DEFINITION yw78b10.r1 Soares_placenta_8to9weeks_2NbHP8to9W Homo sapiens cDNA clone IMAGE:258331 5' similar to gb:M55531 GLUCOSE TRANSPORTER TYPE 5, SMALL INTESTINE (HUMAN);, mRNA sequence.
ACCESSION N40655
VERSION N40655.1 GI:1164252
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 32)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston ,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence starts: 1
High quality sequence stops: 1
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Seq primer: T7
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1. .32
/organism="Homo sapiens"
/db_xref="GDB:3887941"
/db_xref="taxon:9606"
/clone="IMAGE:258331"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/dev_stage="two placentae: one from 8 weeks and another from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 2 a 9 c 12 g 8 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 32;
Best Local Similarity 79.2%; Pred. No. 1.2e+04;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cagtagcagcaacagcatgagacc 25
||| ||||||| |||| | ||||
Db 32 CAGCAGCAGCAGCAGCCTCCGACC 9

RESULT 3
BF346180/c
LOCUS BF346180 55 bp mRNA EST 22-NOV-2000
DEFINITION 602017646F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4153474 5', mRNA sequence.
ACCESSION BF346180
VERSION BF346180.1 GI:11293775
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 55)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM9421 row: k column: 11
High quality sequence stop: 55.
FEATURES Location/Qualifiers
source 1. .55
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4153474"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 4 a 15 c 26 g 10 t
ORIGIN

Query Match 64.0%; Score 16; DB 11; Length 55;
Best Local Similarity 79.2%; Pred. No. 1.3e+04;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 cagtagcagcaacagcatgagacc 25
||| |||||| ||||| |||||
Db 24 CAGCAGCAGCGGCAGCAGCAGACC 1

```
RESULT   4
AA996148
LOCUS       AA996148       52 bp    mRNA            EST       13-APR-1999
DEFINITION  os14b11.s1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1605309 3'
            similar to SW:ASH1_HUMAN P50553 ACHAETE-SCUTE HOMOLOG 1. ;, mRNA
            sequence.
ACCESSION   AA996148
VERSION     AA996148.1  GI:3182637
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 52)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 753   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .52
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1605309"
                     /clone_lib="NCI_CGAP_Lu5"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
                     neuroendocrine lung carcinoid, and was then primed with a
                     Not I - oligo(dT) primer. Double-stranded cDNA was ligated
                     to Eco RI adaptors (Pharmacia), digested with Not I and
                     cloned into the Not I and Eco RI sites of the modified
                     pT7T3 vector.  Library is normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       12 a     19 c     16 g      5 t
ORIGIN


  Query Match            60.8%;  Score 15.2;  DB 10;  Length 52;
  Best Local Similarity  85.0%;  Pred. No. 2.5e+04;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 cagtagcagcaacagcatga 21
          ||| ||||||| |||||| |
Db      5 CAGCAGCAGCAGCATCA 24


RESULT   5
AQ025540
LOCUS       AQ025540       38 bp    DNA             GSS       23-AUG-2000
DEFINITION  EP(X)1614-5prime Drosophila melanogaster EP line Drosophila
            melanogaster genomic Sequence recovered from 5' end of P element,
            DNA sequence.
ACCESSION   AQ025540
VERSION     AQ025540.1  GI:3265892
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 38)
  AUTHORS   Liao,G.-C., Rehm,E.J. and Rubin,G.M.
  TITLE     Insertion site preferences of the P transposable element in
            Drosophila melanogaster
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3347-3351 (2000)
  MEDLINE   20202638
COMMENT     Contact: Gerald Rubin
            Berkeley Drosophila Genome Project
            University of California, Berkeley
            LSA Building, Berkeley, CA 94720-3200, USA
            Fax: 5106439947
            Email: gerry@fruitfly.berkeley.edu
            Sequence recovery method was inverse PCR.

            Sequence orientation is forward strand relative to 5' end of P
            element

            The P element insertion position is base 31 in the 38 bases. This
            insertion position refers to the first base of the 8 base target
            recognition sequence.
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .38
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="Drosophila melanogaster EP line"
                     /note="Inverse PCR was performed on Drosophila
                     melanogaster strains each of which contains a single EP
                     transposable element insertion. (The generation of these
                     insertion strains is described in Rorth P, Szabo K, Bailey
                     A, Laverty T, Rehm J, Rubin GM, Weigmann K, Milan M, Benes
                     V, Ansorge W, Cohen SM. 1998. Systematic gain-of-function
                     genetics in Drosophila. Development 6:1049-1057.) The
                     resultant fragment for each strain was directly sequenced
                     to determine the genomic sequence at the site of
                     insertion. Details of the protocols used can be found at
                     http://fruitfly.berkeley.edu/p_disrupt/inverse_pcr.html."
BASE COUNT       14 a      7 c      11 g      6 t
ORIGIN


  Query Match            60.0%;  Score 15;  DB 13;  Length 38;
  Best Local Similarity  78.3%;  Pred. No. 2.9e+04;
  Matches   18;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 acagtagcagcaacagcatgaga 23
          |||||| |||| | ||  |||||
Db      9 ACAGTATCAGCGAGAGATTGAGA 31


RESULT   6
H66141/c
LOCUS       H66141         49 bp    mRNA            EST       18-OCT-1995
DEFINITION  yu16e05.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:233984 3' similar to gb|M87909|HUMALNE36 Human carcinoma
            cell-derived Alu RNA transcript, (rRNA); gb:M35531 GALACTOSIDE
            2-L-FUCOSYLTRANSFERASE (HUMAN);, mRNA sequence.
ACCESSION   H66141
VERSION     H66141.1  GI:1024881
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 49)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
```

```
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 975
            High quality sequence starts: 1
            High quality sequence stops: 1
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Insert Length: 975   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .49
                     /organism="Homo sapiens"
                     /db_xref="GDB:3787161"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:233984"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT        9 a      9 c     17 g     12 t      2 others
ORIGIN
```

```
  Query Match             60.0%;  Score 15;  DB 11;  Length 49;
  Best Local Similarity   93.8%;  Pred. No. 3e+04;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      10 gcaacagcatgagacc 25
           ||||||| ||||||||
Db      26 GCAACAGNATGAGACC 11
```

```
RESULT   7
C20867
LOCUS       C20867          52 bp    mRNA            EST       23-OCT-1996
DEFINITION  HUMGS0004932 Human adult (K.Okubo) Homo sapiens cDNA 3', mRNA
            sequence.
ACCESSION   C20867
VERSION     C20867.1  GI:1621977
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 52)
  AUTHORS   Okubo,K.
  TITLE     BodyMap; human gene expression database
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Okubo,K.
            Institute for Molecular and Cellular Biol
            Osaka University
            1-3,Yamada-oka, Suita, Osaka Pref. 565, Japan
            Tel: 06-877-5111(ex.3315)
            Email: kousaku@imcb.osaka-u.ac.jp
            Human Gene Signature, 3'-directed cDNA sequence. We are not
            submitting the same cDNA sequence redundantly to DDBJ since 1993.
            For the abundance information of clones with this sequence in this
            library and as well as in other 3'-directed libraries, see '
            http://www.imcb.osaka-u.ac.jp/bodymap'. The sequences of the clones
            represented by this GS sequences is also found there.
FEATURES             Location/Qualifiers
     source          1. .52
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human adult (K.Okubo)"
                     /dev_stage="adult"
BASE COUNT       13 a     18 c     12 g      9 t
ORIGIN
```

```
  Query Match             59.2%;  Score 14.8;  DB 11;  Length 52;
  Best Local Similarity   88.9%;  Pred. No. 3.5e+04;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       7 gcagcaacagcatgagac 24
           ||| |||||| |||||||
Db      27 GCAACAACAGAATGAGAC 44
```

```
RESULT   8
TA180E05P/c
LOCUS       TA180E05P       53 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 180e05, forward sequence,
            genomic survey sequence.
ACCESSION   AL474283
VERSION     AL474283.1  GI:11839571
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 53)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .53
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="180e05"
BASE COUNT        1 a     18 c     11 g     23 t
ORIGIN
```

```
  Query Match             59.2%;  Score 14.8;  DB 13;  Length 53;
  Best Local Similarity   88.9%;  Pred. No. 3.6e+04;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 agcagcaacagcatgaga 23
```

```
          ||||||| ||||| ||||
Db      40 AGCAGCAGCAGCAGGAGA 23


RESULT   9
AZ616824/c
LOCUS       AZ616824       57 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0446L15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0446L15 R, DNA sequence.
ACCESSION   AZ616824
VERSION     AZ616824.1  GI:11739014
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 57)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0446  row: L  column: 15
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 57.
FEATURES             Location/Qualifiers
     source          1. .57
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0446L15"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a      7 c     25 g     25 t
ORIGIN


  Query Match             59.2%;  Score 14.8;  DB 13;  Length 57;
  Best Local Similarity   88.9%;  Pred. No. 3.6e+04;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy       1 acagtagcagcaacagca 18
           |||| ||||||| |||||
Db      30 ACAGCAGCAGCAGCA 13


RESULT  10
AZ779573
LOCUS       AZ779573       24 bp    DNA             GSS       16-FEB-2001
DEFINITION  2M0016K09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0016K09 F, DNA sequence.
ACCESSION   AZ779573
VERSION     AZ779573.1  GI:12910362
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 24)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0016  row: K  column: 09
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 24.
FEATURES             Location/Qualifiers
     source          1. .24
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0016K09"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        8 a      8 c      8 g      0 t
ORIGIN


  Query Match             58.4%;  Score 14.6;  DB 13;  Length 24;
  Best Local Similarity   81.0%;  Pred. No. 3.8e+04;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy      2 cagtagcagcaacagcatgag 22
          ||| ||||||| |||||  ||
Db      2 CAGCAGCAGCAGCAGCAG 22


RESULT  11
AZ404206/c
LOCUS       AZ404206       27 bp    DNA             GSS       03-OCT-2000
DEFINITION  1M0172I20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0172I20 F, DNA sequence.
ACCESSION   AZ404206
VERSION     AZ404206.1  GI:10528219
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 27)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0172  row: I  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 27.
FEATURES             Location/Qualifiers
     source          1..27
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0172I20"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        0 a      9 c      9 g      9 t
ORIGIN


  Query Match             58.4%;  Score 14.6;  DB 13;  Length 27;
  Best Local Similarity   81.0%;  Pred. No. 3.8e+04;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      2 cagtagcagcaacagcatgag 22
          ||| ||||||| |||||  ||
Db     27 CAGCAGCAGCAGCAGCAG 7


RESULT  12
AZ861400
LOCUS       AZ861400       35 bp    DNA             GSS       21-FEB-2001
DEFINITION  2M0167O13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0167O13 R, DNA sequence.
ACCESSION   AZ861400
VERSION     AZ861400.1  GI:13057682
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 35)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0167  row: O  column: 13
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 35.
FEATURES             Location/Qualifiers
     source          1..35
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0167O13"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       12 a     11 c     12 g      0 t
ORIGIN


  Query Match             58.4%;  Score 14.6;  DB 13;  Length 35;
```

```
  Best Local Similarity   81.0%;  Pred. No. 4e+04;
  Matches   17;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       2 cagtagcagcaacagcatgag 22
           ||| ||||||| ||||| ||
Db       3 CAGCAGCAGCAGCAGCAG 23
```

```
RESULT  13
AZ832139/c
LOCUS       AZ832139       40 bp    DNA             GSS       20-FEB-2001
DEFINITION  2M0112P14F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0112P14 F, DNA sequence.
ACCESSION   AZ832139
VERSION     AZ832139.1  GI:13002047
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 40)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0112  row: P  column: 14
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 40.
FEATURES             Location/Qualifiers
     source          1. .40
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0112P14"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        5 a     11 c     10 g     14 t
ORIGIN
```

```
  Query Match             57.6%;  Score 14.4;  DB 13;  Length 40;
  Best Local Similarity   75.0%;  Pred. No. 4.8e+04;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 acagtagcagcaacagcatgagac 24
           | || ||||| ||||| | ||||
Db      33 ATAGCAGCAGGAACAGACTAAGAC 10
```

```
RESULT  14
AZ488023/c
LOCUS       AZ488023       51 bp    DNA             GSS       05-OCT-2000
DEFINITION  1M0318J02F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0318J02 F, DNA sequence.
ACCESSION   AZ488023
VERSION     AZ488023.1  GI:10656316
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 51)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0318  row: J  column: 02
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 51.
FEATURES             Location/Qualifiers
     source          1. .51
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0318J02"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT        3 a      8 c     23 g     17 t
ORIGIN
```

```
  Query Match             57.6%;  Score 14.4;  DB 13;  Length 51;
  Best Local Similarity   75.0%;  Pred. No. 5e+04;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 cagtagcagcaacagcatgagacc 25
           ||| | ||||| || ||| | |||
Db      45 CAGCACCAGCACCACCATCACACC 22


RESULT  15
TA263C05P/c
LOCUS       TA263C05P      53 bp    DNA             GSS       13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 263c05, forward sequence,
            genomic survey sequence.
ACCESSION   AL483802
VERSION     AL483802.1  GI:11849892
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 53)
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES             Location/Qualifiers
     source          1. .53
                     /organism="Trypanosoma brucei"
                     /strain="TREU927"
                     /db_xref="taxon:5691"
                     /clone="263c05"
BASE COUNT        5 a     12 c     14 g     22 t
ORIGIN


  Query Match             57.6%;  Score 14.4;  DB 13;  Length 53;
  Best Local Similarity   75.0%;  Pred. No. 5e+04;
  Matches   18;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 cagtagcagcaacagcatgagacc 25
           ||  ||||| | |||||  |||||
Db      36 CACAAGCAGAAGCAGCACTAGACC 13


Search completed: March  9, 2002, 00:09:14
Job time: 11030 sec
```